

Vonzatkeretlisták helyett polarításos hatásláncsaládok – avagy a \Re ALIS σ függvénye

Alberti Gábor¹, Kilián Imre²

¹ PTE BTK Nyelvtudományi Tanszék
 \Re ALIS Elméleti és Számítógépes Nyelvészeti Kutatócsoport
 albi@btk.pte.hu

² \Re ALIS ESzNyK / PTE TTK Informatika Tanszék
 mindkét cím: 7624 Pécs, Ifjúság útja 6.
 kilian@gamma.ttk.pte.hu

Kivonat: A számítógépes fordításra [5] és más intelligens nyelvfeldolgozási feladatokra [6] irányuló kutatásaink során korábban *esetkeretlisták* formájában tároltuk az igék és más régenek legalapvetőbb vonzatszerkezeti tulajdonságaira vonatkozó információt. Mára megérett a lehetőség a \Re ALIS elmélet nyújtotta dinamikus diskurzus-szemantikai alapokon [10] arra, hogy elméleti nyelvészeti szempontból jóval igényesebb struktúrában (*polarításos hatásláncsaládok* formájában) rögzítsük a rokonítható vonzatszerkezet-változatokat [2], amiből sokkal több információ nyerhető ki a vonzathelyek betöltésére vonatkozóan [1]. Mindezt az „élethossziglani” (‘lifelong’) \Re ALIS keretbe ágyazva tudjuk elhelyezni, egy szövegelemző interpretáló információállapotának részeként [7]. A technológiai oldalon a Prolog nyelv alkalmazásának előnyeiről és mikéntjének fogásairól számolunk be.¹

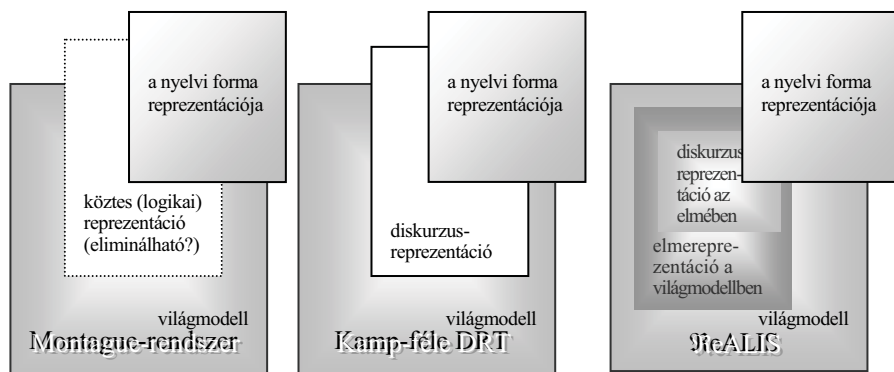
1 A \Re ALIS mint nyelvelméleti keret

A címben szereplő \Re ALIS elméletet a Montague-féle formális szemantika [11] és az abból kisarjadó, kisdiskurzusok jelentésábrázolására kidolgozott (Kamp-féle reprezentacionalista dinamikus szemantikai) DRT elmélet [12] továbbfejlesztéseként szoktuk bemutatni [6, 4].

Az alábbi 1. ábrán összevetjük a megemlített három jelentéselméletnek a nyelvi forma és a világmodell kapcsolatáról felállított modelljét. Ez alkalmat ad arra, hogy felvázoljuk a \Re ALIS kidolgozása melletti döntő metaelméleti érvet. Montague elméletileg igazolta, hogy nincsen szükség a nyelvi forma interpretációja során sem a korábban alkalmazott logikai reprezentációra, sem másfajta közvetítő reprezentációra; de e matematikai tételt nem tudta maradéktalanul átültetni a gyakorlatba. Ez követőinek sem sikerült, sőt egyenesen kialakult egy *reprezentacionalista* irányzat, mely a

¹ A szerzőket e cikk megírásában az OTKA T60595 sz. projektje támogatta. Értékes megjegyzéseikért elsősorban a \Re ALIS ESzNy Kutatócsoport további tagjainak szeretnénk köszönetet mondani: Kleiber Juditnak, Károly Mártonnak és Farkas Juditnak.

közvetítő reprezentáció kiküszöbölhetetlenségét mutatta be nyelvi adatokon, és funkciójára is magyarázatot talált [12]: a *diskurzusábrázolás* önálló szintjeként meghatározva azt. A \Re ALIS-modell képes feloldani a matematikai elmélet és a nyelvi gyakorlat közötti látszólagos ellentmondást egy további tényező, a kettős természetű *interpretálói elme* reprezentációjának a bevonásával.



1. ábra A nyelvi forma és a világmodell kapcsolatának háromféle formális modellje

Az önálló diskurzusreprezentációs szint úgy küszöbölhető ki, ha (nyelvészetiileg releváns) információtartalmát beágyazzuk a világmodellbe, mégpedig a hangfolyamot strukturált diskurzusként felfogó interpretálói elme reprezentációjának részeként. Az interpretálói elmék pedig nyilván részei a világmodellnek, hiszen tartalmuk éppúgy beszéd tárgyát jelent (ki mit tud, hisz, tervez?), mint mondjuk egy lakás részletei, egy állat felépítése, egy város struktúrája. Ami megkülönbözteti a reprezentálandó objektumok körében az elmét az utóbb említett dolgoktól, az az, hogy az elme kettős természetű a reprezentáció folyamatának szempontjából: nemcsak tárgya, de létrehozója is az ábrázolatoknak.

A \Re ALIS-modell ennek az alapgondolatnak a formalizált kidolgozását és nyelvészeti alkalmazását jelenti. A mintegy 40 oldalnyi definíciós állomány (<http://lingua.btk.pte.hu/realispapers>) ismertetésére nyilván ehelyütt nincs mód; most a nyelvtechnológiai relevancia mellett kell érvelnünk. A döntő általános érvünk az, hogy az intelligens nyelvtechnológiai feladatok (például számítógépes jelentésreprezentáció, illetve fordítás) végső soron a kommunikáló emberi elmék lehető legfinomabb modellezését kívánják meg. Idézzünk néhány ezt szemléltető közismert példát: egy másik cikkünkben [7] az alábbi (1) pontban.

Hogy a *miniszter*-re való visszautalás (1a) egy angol fordításban *he* vagy *she* legyen-e, annak eldöntése a szövegkontextus világában való tájékozottságot tesz szükségessé. A magyar szövegben pedig az elnökre utaló *az* névmás jól formáltsága az adott helyen a *diskurzus* szerkesztésének módján múlik: *topikváltás* történik a második mondatban [17].

Az (1b) azt szemlélteti, hogy egy szöveg jólformáltsága (és értelmezési lehetősége) múlhat olyan *ontológiai* ismereteken, mint a bernáthegyi és a kuvasz besorolása a kutyák kategóriájába, a papagájé pedig a madarakéba.

Az (1c) a Szegmentált DRT nevű irányzat [8] egyik iskolapéldája arra, hogy egy szöveg mondatai között (kötőszavak híján is) fel kell tárnunk a *retorikai* relációkat, melyek az *időstruktúrára* is befolyással lehetnek. A nyelvtudomány szempontjából az most a megfigyelni való, hogy ha a második mondatban az első mondat *okát* ismerjük fel, akkor fordíthatjuk angolra Past Perfect idővel; *narratív* retorikai reláció feltételezése mellett viszont nem (ami az erőltetettebb eset). Az *ok* reláció felismerése viszont azt a világismeretet igényli, hogy az *elésnek* a véletlen megbotlás mellett egy akaratlagos ellökés is lehet a kiváltója.

- a. Megpillantottam az elnököt. **Az** viszont nem láthatott meg engem. (2)
- b. Van egy bernáthegyim, egy kuvaszom és egy papagájom.
 - b.1. ... **A két kutya** gyakran felbosszantja **a szegény madarat!**
 - b.2. ... ***A kutya** gyakran felbosszantja **a két szegény madarat!**
- c. Péter elesett. Jancsi csúnyán meglökte.
- d. A kalózvezér **elásatta** a rabolt kincseket.
 - ... **Az emberei** napokig küszködtek **a fagyos földdel**,
 - de még így sem sikerült **kellően mély gödröket** csinálniuk **a hitvány ásóikkal**.

Végül az (1d) diskurzus a cikk központi témáját jelentő *hatásláncsaladók* háttérismeretét szemlélteti, amit egyfajta kiterjesztett *lexikai tudás* részének tartunk. Az *elásat* igealak explicit módon csak egy Okozó szereplő (a kalózvezér) és egy Páciens jellegű szereplő (a kincsek) relációját mutatja be. A második mondat kézenfekvő értelmezése viszont feltételezi az alábbi hatáslánc felidézését (ami a jelentésábrázolás vagy -kivonatolás során is releváns): az Okozó *közvetlen* hatást (mondjuk egy kiadott parancs révén) az Ágensre gyakorol (az embereire), akik Eszközeik (a hitvány áók) segítségével (fagyos) földdarabokat (szintén Páciens?) ásnak ki eredeti helyükről (Kezdőpont?), egy vagy több gödröt létrehozva; a kincsek végül is idekerülnek az Okozó akaratából. Az említett szereptípusok a *thematikus szerepek elméletének* [15] az eszköztárából származnak, amelynek *absztrakt szerephierarchiákra* épülő továbbfejlesztett modelljét [2, 4] vesszük alapul e projektben.

Azt az álláspontot képviseljük tehát, hogy az intelligens nyelvtudományi feladatok olyan modellezését kívánják meg a kommunikációban álló emberi elméknek, ami hatalmas (és dinamikusan fejleszthető) kulturális/enciklopédikus, ontológiai és erősen kiterjesztett lexikai információs bázison alapul. A matematikailag formalizált *ReALIS*-modell (*Reciprocal and Lifelong Interpretation System*) éppen ezt kínálja, „kölcsonös” és „élethossziglani” jellegéből adódóan: a kölcsönösség a kommunikációs interakció megragadását biztosítja, az élethossziglaniság pedig hatalmas, folyamatosan bővíthető adatbázisokat jelent. Egyszerűen fogalmazva, a számítógépes jelentéskivonatolás és fordítás jövőjét abban látjuk, hogy a gép egykor úgy fog majd működni, mintha a megfigyelő/fordító (interpretáló) ember tevékenységét szimulálná.

A *ReALIS* interpretáló modelljének *szimultán rekurzív* definícióját röviden és informálisan úgy vázolhatjuk fel, mint egy élethossziglani folyamat leírását, amelynek során egy kezdetben (a születés idealizált pillanatában) strukturálatlan referenshalmozon négy reláció terjeszkedik ki, mintegy leképezve a környező világ hatásait. Az α *horgonyzófüggvény* az ugyanazon szereplőre mutató referenseket társítja egymással (az interpretáció során leginkább a régenszekhez megtalált vonzatok és a határozott főnévi kifejezésekhez megtalált antecedensek alapján). Egy interpretálói elmében α

relációban áll például rengeteg olyan referens, ami W. A. Mozarthra vonatkozó információ formulájában szerepel, a híres zeneszerzőre utalva. A λ szintfüggvény a klaszikus szemantikából ismert konstans-változó megkülönböztetés általánosításaképpen fogható fel, e kételemű reláció helyett egy gazdag részenrendezési struktúra osztályaiba („világocskákba”) sorolva a külvilág entitásaira utaló referenseket és az interpretálói hiedelmek, vágyak, feltételezések, szándékok és tervek „fiktív” referenseit. A κ kurzor az interpretáló koncentrált figyelmét kívánja modellezni, pillanatonként változó módon kijelölve idő-, tér-, topik- és eseményreferenseket. Végül a σ eventuális függvény feladata „összerakni” a referensekből azokat a formulákat, amelyek megragadni hivatottak az eseményeket és állapotokat (azaz „eventualításokat”). Ez teszi lehetővé a *robustusságot*: külön-külön is információt adnak a referensek, nemcsak egyben létezik egy-egy propozícióformula, mint a korábbi elméleteknél; így a hiányos mondatokhoz is rendelhető valamilyen reprezentáció.

Az alábbi (2) eventuális formula például nagyjából egy olyasféle eseményt ($e_{\text{elásatás}}$) ír le a DRT „nyelvén” [12], hogy Long John Silver elásatott velem valamilyen kincseket egy szigeten egy bizonyos időpontban ($t_{2010-09-11}$).

$$e_{\text{elásatás}} : p_{\text{elásat}} t_{2010-09-11} r_{\text{LongJohnSilver}} r_{\text{kincsek}} r_{\text{én}} r_{\text{sziget}} \quad (3)$$

Ezt a ReALIS (kétváltozós) σ függvénye úgy ragadja meg, hogy egy referenshez (pl. $e_{\text{elásatás}}$) (ami ezáltal *eventuális referensnek* osztályozódik) és egy szerepparaméterhez hozzárendel egy predikátumot (pl. $p_{\text{elásat}}$), egy időreferenst (t_x) és 0 vagy több argumentumreferenst (r_y). Ez idáig jóformán csak technikai újítás, a lényeg most jön: a szerepparaméter értékvektora választ néhány szereplőt egy *polaritásos hatáslánccsaládból*, beállítva az alábbi 1. táblázat bal oldalán felsorolt összes tényezőt a jobb oldali lehetőségek szerint. A következőkben bemutatjuk, hogy ezzel a megközelítéssel milyen széles körben meg tudjuk ragadni az ige (vagy más régens) körül kibontakozó mondat(rész) grammatikai jellemzőit, lehetőségeit, járulékos hangtani és jelentéstani tulajdonságait. Mindenekelőtt azonban megismerkedünk a polaritásos hatáslánccsaládokkal.

1. táblázat: Az argumentumszerepet meghatározó paraméterek és értékeik.

HATÓKÖRI SORREND	1, 2, 3, 4, 5, ...
HATÁSLÁNCCSALÁDBELI SZEREP(CÍMKE)	pl. Ág, Pác, Pác', Pác'', Ok, Esz, Idő, Tér
ESETPROMINENCIA [2]	centrális (\rightarrow), nem centr. vonzat (\bullet), szabad hat. (\circ)
INFORMÁCIÓS SZEREP [14]	$\{T, K\}^{*^*} \{F, Q\}^{*^*} (M)^C$
REFERENCIALITÁSI FOKOZAT [1]	+hat > [-hat, +spec] > [-spec, +ref] > [-ref, +exp] > \emptyset
BESZÉDAKTUSBELI RÉSZVÉTEL (SAP) [6]	1Sg > 2Sg > ... > 3Pl

2 A polaritásos hatáslánccsaládok

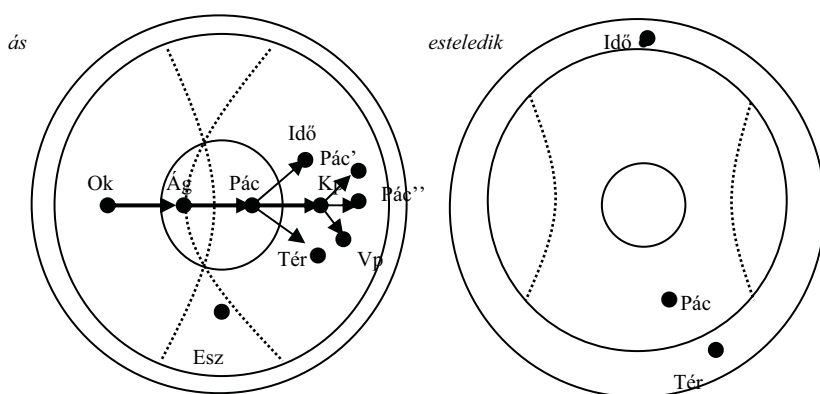
Az alábbi 2. táblázat thematikus szerepekkel megadott vonzatszerkezet-változatokat mutat, a termékeny *ás* és a nem igazán termékeny *esteledik* igei családjából.

Hogy a fenti 1. táblázatban *centrálisnak* nevezett alanyi és tárgyi grammatikai funkciót mely vonzatszerpekhez társíthatjuk, azt univerzális szabályok korlátozzák

[2, 3], az adott családra jellemző karakter rögzítésén túl, amit az alábbi 2. ábra mutat be vizuálisan. A lényeg az, hogy a pontsorral lehatárolt bal oldali körszeletben vannak a potenciális tranzitív alanyok („ágensi karakter”), a jobb oldaliban pedig a tárgyak („páciensi karakter”) – ez adja a *polaritások* jelleget; továbbá olyan ⟨Alany, Tárgy⟩ vonzatkeret nem lehetséges, amelyben a Tárgy felől mutatna a hatásirány nyíla az Alany felé – ez ugyanis megsértené a hatáslánc elvét.

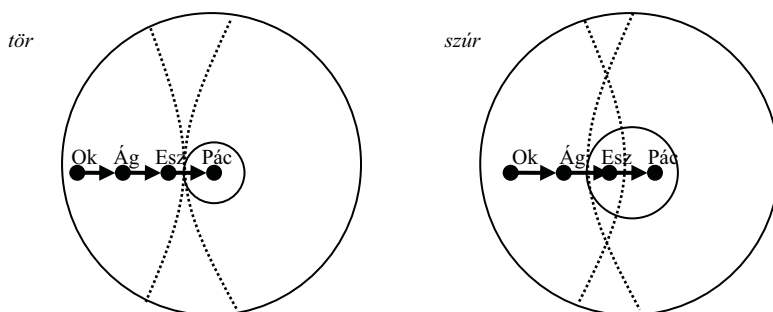
2. táblázat: Két magyar igező néhány vonzatkerete.

CENTRÁLIS ESETKERET	<i>esteledik / ás</i>
⟨ ⟩ (Páciens, Idő, Tér)	1. a. Esteledett. / b. Ránk esteledett (8-kor a tóparton).
⟨Ágens, Páciens⟩	2. A kalózok (kétségbeesetten) ásták az agyagos földet.
⟨Ágens, Tér/Idő⟩	3. a-b. A kalózok felásták a szigetet / végigásták a hétvégét.
⟨Ágens, Kezdőpont⟩	4. A kalózok ástak egy mély gödröt.
⟨Ágens, Végpont⟩	5. A kalózok ástak egy sírt.
⟨Ágens, Páciens' / Pác.'⟩	6. a-b. A kalózok elástak / kiástak egy értékes kincset.
⟨Ágens, Eszköz⟩	7. – (* A kalózok kétségbeesetten ásták a hitvány ásóikat.)
⟨Okozó, Pác' / Pác''' / Pác'⟩	8. a-b. a-b. A vezér elásatta / kiásatta a kincset.
⟨Okozó, Ágens⟩	9. ?? A vezér (álló nap) ásatta az embereit.

2. ábra. Két magyar igező (részleges) polaritások hatásláncsaládjá: *esteledik* és *ás*.

Az imént meghivatkozott cikkeinkben részletesen ismertetjük, hogy hogyan lehet néhány nyelvi adatból meghatározni egy-egy igező esetében a *polaritások hatásláncsaládot*. Feltételezzük, hogy az „ideális interpretáló” is hasonló módon figyeli meg a hatásirányt, továbbá az ágensi/páciensi karaktert a vonzatszerpek esetében; ezek alapján aztán felépíti magában a polaritások hatásláncsaládot, ami a továbbiakban meghatározza, hogy milyen vonzatszerkezet-változatokat ítél majd jól formálnak (ellentmondó adatok esetén esetleg módosítva a hatásláncsalád struktúráját). A nyelvtechnológiai alkalmazás ezek után kézenfekvő, összhangban az interpretáló szimulálásának elvével: vonzatkeretlisták helyett polaritások hatásláncsaládok formájában tároljuk a releváns lexikai információt.

Újdonság értékű megállapításunk [2, 3] az, hogy a lehetséges vonzatszerkezet-változatok nem (közvetlenül) a thematikus karakterből adódnak, hanem a *hatáslánc*-irányon kívül egyrészt a polaritásos hatáslánccsaládok *primitív magjából*, amit a 2-3. ábrán a legbelső kör jelöl, másrészt a (pontsorral jelölt ívek által lehatárolt bal / jobb oldali) *ágensi*, illetve *páciensi pólusból*. E négy tényező részben véletlenszerűen alakul ki a nyelv története során egy-egy igető esetében: ebből adódik az alábbi 3. táblázatban szemléltetett különbség a *tör* és a *szúr* lehetséges vonzatkeret-változatai között, miközben a vonzatszerpek thematikus karakterében erőltetett lenne bármi különbséget feltételezni (3. ábra). Arról van szó, hogy a *szúrás* Eszköze tranzitív alanyként és tárgyként egyaránt szerepelhet, míg a *törés* Eszköze tárgyként nem.



3. ábra. Két hasonló magyar igető (részleges) polaritásos hatáslánccsaládja: *tör* és *szúr*.

3. táblázat: A *tör* és a *szúr* vonzatkeret-változatai.

<i>tör</i>	CENTRÁLIS ESETKERET	<i>szúr</i>
Betört egy ablak. <i>primitív mag!</i>	⟨Páciens⟩	–
Ez a kalapács még vastagabb ablakot is betörne.	⟨Eszköz, Páciens⟩	Egy szög megszurta a kezemet. <i>primitív mag!</i>
–	⟨Ágens, Eszköz⟩	Péter beleszúrt egy szöget az abroncsba.
Péter betörte az ablakot egy kalapáccsal.	⟨Ágens, Páciens⟩	Péter kiszúrta az abroncsot egy szöggel.
–	⟨Okozó, Eszköz⟩	Mari beleszúratott egy szöget az abroncsba.
Mari betörtette az ablakot (Péterrel / egy kalapáccsal).	⟨Okozó, Páciens⟩	Mari kiszúratta az abroncsot (Péterrel / egy szöggel).
–	⟨Okozó, Ágens⟩	–

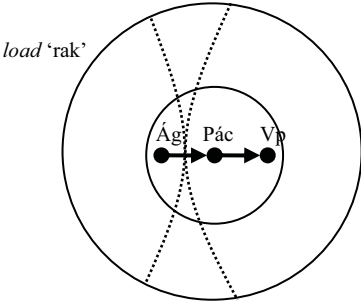
A *ReALIS* modelljében tehát a hagyományos thematikus szerepek [15] csupán címkéként szerepelnek, szemantikai karakterük a polaritásos hatáslánccsalád struktúrájában betöltött pozíciójuk révén definiáltak.

Térjünk vissza az elméleti kitérő után a 2. ábrához! Az *esteledik* primitív magja üres, amint azt a T2.1.a. példamondat mutatja a fenti 2. táblázatban. Az Idő és a Tér szabad határozói (l. a külső gyűrűben!) csatlakozhatnak ehhez az igetőhöz, illetve vonzatként egy sajátos Páciens (T2.1.b.), nem alanyi vagy tárgyi (azaz centrális) funkcióban, hanem *-rA* ragos alakban. Az *esteledik* hatáslánccsaládja ennyi.

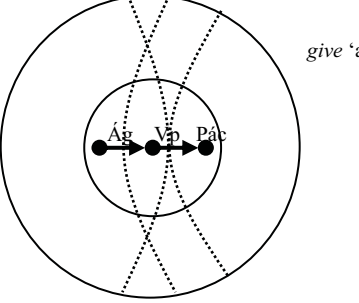
Az *ás*-é viszont annál gazdagabban burjánzik! A legszűkebb jelentés, a primitív mag földdarabok mozgatását írja le (T2.2.). E földmozgató tevékenység kumulálódhat alternatív módokon: behatottá téve egy földterületet (T2.3.a.), egy időszakot (T2.3.b.), vagy *gödört* eredményezve a Kezdőpontból távozó földdarabok helyén (T2.4.). A gödörásásból kiindulva is többféle jelentésbővülést tapasztalhatunk, újabb tranzitív vonzatkeret-változatok formájában: a gödör szolgálhat sírként (T2.5.), vagy kincsek rejtékhelyeként – ez utóbbi esetben az *ásás* tevékenysége kétféle célt is szolgálhat (T2.6.a-b.). Az *ás* polaritásos hatásláncsaládjában a pontsorívek helyzete megmutatja, hogy az *ásás* Eszköze alanyként / tárgyként nem fejezhető ki (T2.7.), az Ágens tárgyi kifejezése pedig „határeset” (T2.9.). További jelentésbővülést egy Okozó bevonásával érhetünk el (T2.8.a-b.; l. még (1d) és (2)), ami mellett tárgyként bármely korábban szóba került tárgyat is felvehetünk (pl. *sírt ásat*).

A szakaszban tárgyalt vonzatemélet univerzális. Ennek illusztrálása végett most (kommentárok nélkül) bemutatjuk az egytárgyas és a kéttárgyas angol igetípus tárgy-alternációjának és passzivizálásának lehetőségeit, illetve az ezeket meghatározó polaritásos hatásláncsaládokat.

4. táblázat/ábra. Két angol igető (részleges) polaritásos hatásláncsaládjá.



load 'rak'



give 'ad'

load 'rak'	CENTRÁLIS ESETKERET	give 'ad'
He loaded hay onto the wagon. ő rak-Past széna -rA a kocsi 'Szénát rakott a kocsira.'	⟨Ágens, Páciens⟩	He gave a book to Mary. ő adott egy könyv Mari 'Adott egy könyvet Marinak.'
He loaded the wagon with hay. ő rak-Past a kocsi -vAl széna 'Megrakta a kocsit szénával.'	⟨Ágens, Végpont⟩	—
—	⟨Ágens, Végpont, Páciens⟩	He gave Mary a book. ő adott Mari egy könyv 'Adott Marinak egy könyvet.'
The hay was loaded onto wagons a széna volt rak-PastP -rA kocsi-PI 'A szénát kocsikra rakták.'	⟨Páciens⟩	The book was given to Mary. egy könyv volt ad-PastP -nAk Mari 'A könyvet odaadták Marinak.'
—	⟨Végpont, Páciens⟩	Mary was given a book. Mari volt ad-PastP egy könyv 'Marinak adtak egy könyvet.'
The wagon was loaded with hay. a kocsi volt rak-PastP -vAl széna 'A kocsit megrakták szénával.'	⟨Végpont⟩	—

3 A σ függvény grammatikai paramétervektora

Mint az 1. szakasz végén leszögeztük, a $\Re\text{ALIS}$ modelljében egy esemény vagy állapot reprezentációját a σ eventuális függvény szervezi meg, hozzárendelve egy eventuális referenshez a predikátumot képviselő referenst és az argumentumokat képviselő referenseket (valamint egy időreferenst).

Ami a *predikátum* kijelölését illeti, arra az előnyös megközelítésre szeretnénk felhívni a figyelmet, hogy amennyiben az interpretáló például a fenti 2. táblázat valamely mondatában szembesül az *ás* ige-tővel, nem kell egyből döntenie, hogy melyik vonzatkeret-változatról van szó, hanem annyit kell megállapítania, hogy a 2. ábrán bemutatott bal oldali családot „hívja elő” az *ás* hangalak, rögzítve valamiféle alapjelentést (a potenciális bővítésekkel együtt). Az adott vonzatkeret-változat meghatározása azután egy specifikáló procedúrára van bízva.

E specifikáció az argumentumreferensek *paramétervektorának* beállításával folytatódik (ami technikailag a kétváltozós σ függvény egyik argumentumhelyére kerül, az eventuális referens mellé). Ennek egyik fontos eleme, amint ezt az 1. szakasz végén közölt 1. táblázat 2-3. sora is mutatja, *szerepcímkek* segítségével kiválasztani a polaritások hatásláncsaládból a predikátummal társítandó argumentumokat, és specifikálni, hogy melyikük kapjon *centrális* funkciót, azaz alanyi / tárgyi megjelenítést.

Az alábbi 5.a-b. táblázat egyetlen vonzatkeret-választást szemléltet a számtalan megengedett változatból: az Okozónak és az elásandó kincsre utaló Páciens' címkejű szereplőnek biztosítunk centrális grammatikai funkciót; a hatáslánc elvéből adódóan az már determinált, hogy az előbbi alanyként, az utóbbi tárgyként jelenítendő meg.

5. táblázat: Két argumentumkiosztás az *ás* predikátumreferens mellett.

a. <i>ás</i>	hatókör	szerepcímke	esetprom.	inf. szerep	ref. fokozat	SAP
+ref	1	Ok	→	T	+hat	3Sg
(+ref) _F	2	Tér	○	F	-hat, +ref	3Sg
+ref	3	Pác'	→	C	+hat	3Pl
+ref	4	Ág	●	C	+hat	1Sg

A 'kalózvezér egy ''lakatlan szigeten ásatta el velem a rabolt kincseket.

[?]A 'kalózvezér egy ''lakatlan szigeten ásatta velem el a rabolt kincseket.

[?]A 'kalózvezér egy ''lakatlan szigeten ásatta el a rabolt kincseket velem.

^{??}A 'kalózvezér egy ''lakatlan szigeten ásatta velem a rabolt kincseket el.

^{??}A 'kalózvezér egy ''lakatlan szigeten ásatta a rabolt kincseket velem el.

^{???}A 'kalózvezér egy ''lakatlan szigeten ásatta a rabolt kincseket el velem.

b. <i>ás</i>	hatókör	szerepcímke	esetprom.	inf. szerep	ref. fokozat	SAP
(+ref) _F	1	Tér	○	F	+hat	3Sg
(+ref) _F	2	Ag	●		-hat, +ref	3Sg
+ref	3	Ok	→	C	+hat	3Sg
+ref	4	Pác'	→	C	+hat	3Sg

^{???}Madeirán ásatta el a kalózvezér a rabolt kincseket egy ''tússzal.

[?]Madeirán ásatta el egy ''tússzal a kalózvezér a rabolt kincseket.

^{??}Madeirán ásatta egy ''tússzal el a kalózvezér a rabolt kincseket.

^{???}Madeirán ásatta el a kalózvezér egy ''tússzal a rabolt kincseket.

Az 1. táblázat azt is megmutatja, hogy egy mondatba kerülő argumentum formai tulajdonságait számos más paraméter értékelése is befolyásolja, a tematikus szerepén és a grammatikai funkcióján kívül. Ezeket érdemes egyetlen paramétervektorban

egyesíteni, mivel ily módon minden nyelvészeti megszorítást e vektor (részben nyelvspecifikus, részben univerzális) jólformáltsági feltételeként fogalmazhatunk meg, illetve építhetünk be a gépi nyelvfeldolgozó rendszereinkbe. A fordítás problémája innen nézve annak kérdése, hogy egy adott paramétervektor egyaránt jól formált-e a forrás- és a célnyelvben; ha pedig nem, akkor a forrásnyelvi paramétervektorhoz milyen jól formált célnyelvi paramétervektor áll a legközelebb. A kérdéskör nyelvészeti oldaláról a [11] cikk nyújt részletes tájékoztatást; ehelyütt csak szemléltetjük a lehetőségeket az 5.a-b. táblázat alapján.

A magyarban az argumentumok *hatóköri* sorrendje független a tematikus szerepüktől és a grammatikai funkciójuktól. A T5.a. verzióban ilyen hatóköri sorrendet tekintünk: Okozó > Tér > Páciens' > Ágens; a T5.b. verzióban pedig ilyet: Tér > Ágens > Okozó > Páciens'.

Az *információs szerkezeti* szereposztás viszont erősen függ a hatóköri sorrendtől, bár a korábbi Topikfélék > Kvantorfélék > Fókusz sorrendet az újabb kutatások [14] „liberalizálták” (lényegében az 1. táblázat 4. sorában megadott módon). A T5.a. verzióban az Okozó topik funkciót kap, a Tér argumentuma pedig fókusz funkciót. A T5.b-ben egy <Tér, Ágens> tükörfókusz konstrukciót szemléltetünk. A 'C' ('komplementum') jelölés arra utal, hogy egy argumentum nem kapott semmilyen kitüntetett szerepet az információs szerkezetben. Az 'M' a generatív szakirodalomban sokat tanulmányozott igemódosítói / igeívői pozícióra utal. Az alábbi 6. táblázatban felidézzük a logikai jelentéstöbbletet adó információs szerepek jelentésrendszerét:

6. táblázat: A magyar operátorjelentések rendszere.

($R_n = R \setminus R_m$, ahol R_m : a megemlített szereplők, R : minden olyan szereplő, amely eljátszhatta volna a megemlített szereplők által eljátszott szerepet)

	$P(x)$	$\neg P(x)$
$\exists x \in R_n$	Q_{is} : <i>is</i> kvantor <i>Meglep, hogy a nővéremet is meghívtad.</i>	K : kontrasztív topik <i>A nővéremet bezzeg meghívtad!</i>
$\forall x \in R_n$	Q_{mind} : <i>mind</i> kvantor <i>Meghívhattál volna mindannyiunkat!</i>	F : azonosító–kirekesztő fókusz <i>Bánt, hogy csak Annát és Beát hívtad meg.</i>

Az argumentumok *referencialitási fokozatát* korlátozhatja a tematikus szerep (l. [1]: pl. **Alakult az énekkar a klubban*), de a korlátokat semlegesítheti az információs szerep (pl. egy fókusz funkció: *A klubban alakult az énekkar*). Az ezzel kapcsolatos nyelvspecifikus, illetve univerzális tudás is beépíthető a σ függvény paramétervektorának jólformáltsági feltételeibe. Új ötlet, hogy a szám / személy értéket is vegyük figyelembe (SAP: 'speech act prominence'), mivel korrelációt mutat például a topikválasztással [10].

Ahogy az 5.a-b. táblázatok alatt megadtuk, a paramétervektor teljes specifikálását követően preferenciasorrendbe álló szórendi és hangsúlyozási változatok egész listája adódik. Gyakran több kifogástalan szórend is van, például abból adódóan, hogy a magyar kvantor nem feltétlenül áll az ige elé megmutatni a hatóköri elsőbbségét. A 6. táblázat megfelelő példamondata így is elhangozhatna: *Meglep, hogy meghívtad a nővéremet is.*

4. Implementáció: egy megvalósítás sarokkövei

A ReALIS szintaxisfelbontó hozzáállása – a szemantikai relációk felépítését nem a szintaxistól függetlenül, hanem a kétirányú hozzáállást keverve, együtt, közösen végezzük, és a szemantikai információkat is a lexikon elemeihez kötjük –, komoly felkészültséget és gondos tervezést követel. Ilyen bonyolultsági fokozatot csakis valamiféle szimbolikus programozási nyelv, a „mesterséges intelligencia” programnyelvek valamelyikének választásával érhetünk el. Az előzetes tapasztalatok alapján mindegyre a Prolog programozási nyelv valamelyik dialektusa alkalmas.

Bemenet és kimenet. Egy program tervezésénél az első lépés: a bemenet és kimenet specifikációja, vagyis a program által kiértékelt ún. „fekete doboz” függvény értelmezési tartományának és értékkészletének a megadása. Ez kiindulópontul szolgál olyan értelemben is, hogy egy részleges megvalósítás (pl. többdimenziós halmazok esetén) ennek részalmazát, vagy függvényösszetétel esetén a teljes feldolgozási függvény valamilyen összetevőjét számolja ki.

A ReALIS-megvalósítás célkitűzése a szöveg és a diskurzusreprezentáció közötti reláció kiszámítása. Ez (Prolog-szerű értelmezésben) mindkét irányú kapcsolatot jelenti. Ha a szöveg adott, akkor a program azt a reprezentációs kifejezést számítja ki, amely az adott logikai rendszerben és az interpretáló belső információállapotát leíró tudásbázisban (ontológiában) kiértékelhető, bizonyítható vagy hozzávehető a tudásbázishoz. Az ellenkező irányban: ha a tudáskezelő összetevő által (pl. egy kérdésre adott válaszként) egy logikai kifejezést kapunk, akkor végeredményben a szöveg képét állítja elő.

Nemdeterminizmus. Egyes helyzetekben nem eldönthető, hogy egy elemzési folyamatot milyen irányban érdemes folytatni, illetve jó okkal többféle irányban is folytatható lenne (pl. a *vár* szó előfordulása esetén, a homonímia miatt). Bizonyos szó vagy mondat elolvasásakor tehát még nem állapítható meg, hogy melyik értelmezés a helyes. A mondat vagy a diskurzus további elemeinek az elolvasása a helyzetet általában egyértelműsíti. Az ilyen helyzetek kezelésére egyértelműen a *nemdeterminisztikus* megoldások javasolhatók. A program eljárásai kiszámítják a relációk összes lehetséges értékét, amely értékalmazból a hívó eljárások kiválasztják a számukra megfelelőket, és csak azok figyelembevételével állítják elő a saját eredményeiket.

Tárgygépmodell. A mi esetünkben a „tárgygépmodell” kifejezés talán tárgymodellként is felfogható, hiszen nem valamilyen konkrét processzorlapka regiszter- és utasításkészletéről van szó, hanem az elsőrendű logika valamelyik részosztályának használatáról és a nyelvi fogalmak elsőrendű logikába történő leképezéséről. Ez lehet maga a *Horn-klózek* részosztálya is, de előfordulhat az is, hogy ennél bővebb osztályt kell választanunk (pl. következményoldali diszjunkcióval, modális operátorokkal kiegészítve). Ha a *Horn-klózek* szintjén maradunk is, külön szerencsének kell felfognunk, ha az egész a tiszta Prolog következtetési mechanizmusával meghajtható – a feladat egyébként esetleg a Prologtól különböző rezolúciós stratégia alkalmazását is igényelheti [13].

Adatszerkezetek leírása elsőrendű logikában. Az elsőrendű logika alapvetően típusatlan. Típusfogalom mégis alkalmazható úgy, hogy az alkalmazható függvénykifejezések körére megkötéseket teszünk. A Prolog-szakirodalomban a fogalom a vál-

tozók lekötési állapotát leíró ún. móddeklarációk kiterjesztéseként vált ismertté [16]. Bár típusos Prolog-megvalósítás létezik [9], teljes, tiszta Prologhoz használható adatszerkezet-megadó nyelvet eddig mégsem rögzítettek, és erre vonatkozó megvalósításról sem számol be senki. A következőkben intuitíven érthető és követhető példákat mutatunk be adatszerkezet-megadásra. A leképezés egyik buktatója a *másodrendű szerkezetek megvalósítása elsőrendű eszközökkel*. Erre két lehetőség adódik:

Reifikáció, vagyis a predikátumszimbólumot adatként kezeljük, így akár változó is felveheti értékül a futás során. Ennek feltétele, hogy a predikátumot ne kelljen predikátumként kiértékelni, vagy ha mégis, akkor a kiértékelés a tételbizonyítási folyamat jól megfogható helyzetében történjék, amikor a predikátumot logikán túli eszközökkel felépítjük, és azon túl már maga is részt vehet a tételbizonyítás folyamatában.

Logikán kívüli eszközök közvetlen alkalmazása: a Prolog-rendszerekben erre a célra szolgálnak a különböző dinamikus kifejezés-felépítő eljárások (pl. $a = . / 2$). A megoldás hátránya, hogy nem eléggé deklaratív: a lexikális tudás ábrázolása a befoglaló logikai rendszerre (Prolog) nézve specifikus eszközöket használ.

Azonosítás, referensek. *Referenseket* olyan változókkal jelölünk, amelyek kezdetben ismeretlenek, később a referens típusától függő értéket vehetnek fel. *Objektumreferensek* azok, amelyek a világ egy egyéniségét (individuumát, pl. Petőfi Sándor) címzik meg. Ezek a tudásbázis valamely osztályának példányai, az ilyen osztályokat a természetes tulajdonságértékeik nem azonosítják: az adatbázis (pl. telefonkönyv) egy másik Petőfi Sándora még akkor sem egyezik meg a költővel, ha egyébként a többi adatuk is véletlenül megegyezne. Az ettől eltérő adatszerkezeteket *adat-típusnak* nevezzük, ezeket az értékeik egyértelműen azonosítják. Efféle típusok például az időreferensek és a címek.

Időreferensek használata időértékekkel lekötődő változóval lehetséges. Az időreferensek leírják a múltban, jelenben és jövőben történő eseményeket, tetszőleges pontossággal, időpontokat és időszakaszokat egyaránt. A *past (date (1-0-0))* időreferens például a "tavaly ilyenkor" időmeghatározást jelenti nap pontossággal.

Predikátumreferensek az elsőrendű logikában csak *reifikációval* használhatók. A *REALIS* predikátumreferensei a predikátummintát adják meg, amelyet a $\wedge/2$ kapcsolóval kapcsolunk a minta szabad változóihoz. Például az $AG \wedge PAT \wedge TIME \wedge dig (AG, PAT, TIME)$ predikátumreferens az „elás” háromváltozójú (ágens, páciens, időpont) alakját jelöli. Megjegyezzük, hogy a predikátumreferenseken keresztül kötjük a mondat logikai alakját az interpretáló belső állapotát jelképező ontológiához.

Az *eseményreferensek* a predikátummintákból az ismert logikai kapcsolók útján létrehozott kifejezéseket jelölik. A predikátumokból alkotott literálisok, illetve az ilyenekből logikai kapcsolókkal képzett kifejezések ebben a helyzetben adatként viselkednek, melyek pontos szerepe (ontológiabővítés, bizonyítás stb.) a programkörnyezetből derül ki.

Következtetési és relációs tárgymodell. A *relációs szemléletmód* alatt azt értjük, amikor a felismerést (és a generálást is) egyetlen relációban valósítjuk meg. Ennek a relációnak az egyik paramétere a szóban forgó diskurzus belső reprezentációja, a másik paramétere a mondat szöveges alakja. Ez a szemléletmód a Prologot ismerők számára kézenfekvőbbnek tűnhet; hátránya az, hogy a mondatot szekvenciális fájll-

szerűen olvassa, amelyben a többszöri oda-vissza tekerések mindenképpen hatékonyságvesztést és az ábrázolástól idegen megközelítést jelentenek.

A következtetésre alapuló ábrázolásmód a felismerés feladatát hangsúlyozza, és párhuzamba állítja a Horn-klózonokon alkalmazott következtetési folyamattal. Vagyis: a Horn-féle következtetési háló forrásai a tényállítások, amelyek a mi esetünkben a beolvasott diskurzus elemei (szavai, morfémai, betűi). Ezeken végzünk felismerési–következtetési műveleteket, amelyek valamilyen végkövetkeztetésben (célállításban) érnek véget. A végkövetkeztetés célszerűen egy olyan állítás lehet, amely paraméterként a teljes diskurzus szerkezetének valamilyen Prolog leképezését adja eredményül. A „lebegő” nyelvtani elemek és referensek, valamint azok összekapcsolódása a következtetési megoldásban sokkal természetesebb.

A következtetési hálón célszerű rétegeket (vágatokat) értelmezni (I-IV.):

I. A szóalaktani elemzés szintje. Ez a beolvasott karaktersorozaton dolgozik, és ebből előállít egy szó- (morféma-) listát, valamint az egyes elemek diskurzuson, mondaton, szón belüli indexeit. A továbbiakban feltételezzük, hogy ez a lépés már problémamentesen működik, a szógyököket és szóösszetételeket – ahol kellett, azonosította, a hozzá kapcsolódó igekötő-, képző- és ragrendszer felépítését, így a nyelvtani eseteket és egyeztetéseket felderítette. Az elemzés eredményét az alábbihoz hasonló tényállításokban rögzítjük:

word(1,noun('Petra',proper,nom,sing-3)).

II. A szószerkezet-átmenet szintje. Ez gyakorlatilag a régens–vonzat, az alaptag-adjunktum, valamint az anafora–előzmény nyelvtani viszonyokat deríti fel, a lexikai leírásokban megjelenő követel / kínál viszonyokra alapozva.

III. Eventualitások meghatározása a polaritásos hatásláncmodell alapján, következtetésre alapuló ábrázolásban. Az egyes szófajok előfordulása implikálja a szemantikai predikátum fennállását.

```
sigma(AG^PAT^TIME^INST^CAUS^stabWithInit(
    AG,PAT,TIME,INST,CAUS)) :-
    noun(CAUS,_,nom,AGR), noun(PAT,_,acc,_-3),
    verb(caus('szúr'),_PREFIX,decl,VTIME,AGR),
    noun(INST,common,inst,_-3),
    time2verb(TIME,VTIME).
```

IV. Eventualitások meghatározása a hatásláncmodell alapján, relációs ábrázolásban. Az egyes szófajok előfordulása relációban áll a szemantikai predikátummal.

```
sigma(AG^PAT^TIME^INST^CAUS^stabWithInit(
    AG,PAT,TIME,INST,CAUS)),
noun(CAUS,_,nom,AGR), noun(PAT,_,acc,_-3),
verb(caus('szúr'),_PREFIX,decl,VTIME,AGR),
[noun(INST,common,inst,_-3)]) :-
time2verb(TIME,VTIME).
```

A fenti példában a „Mari kiszúratta az abroncsot egy szöggel” példamondat *szúr* igéjének predikátumkiszámító részletét láthatjuk, melyet az ontológia szintjén a

`stabWithInit/5` predikátummal ábrázolunk. Ez a műveletessel és eszközzel történő *kiszűrást* fejezi ki, és az általános AG (ágens), PAT (páciens), TIME (idő) paraméterek mellett még hivatkozik az INST (eszköz) és a CAUS (kezdeményező) paraméterekre is. A `time2verb` feltétel a `stabWithInit/5` predikátum általános időreferense és az igeidő közötti kapcsolatot számítja ki. A következtetési modellben a `stabWithInit` predikátum az állítás következményrészében reifikálva, míg az egyes szóalkotóelemekre hivatkozás a feltételrészében szerepel. Ez utóbbiak a relációs modellben az állítás további paramétereiként láthatók.

A bemutatott lexikonrészletből létrejött program egyrészt meglepően egyszerű – lényegében a cikkben említett egyes példaigék (*ás, szűr*) hatásláncmodell-változatainak rögzítését tartalmazza. A tesztsorral való meghajtás eredménye egyfelől az elvárásokat hozta. Másfelől pedig rámutatott: további szemantikai vagy környezeti információk feldolgozása nélkül a helyzet nem egyértelmű, ami a nemdeterminisztikus eredmények túlbujánzásához vezet. Például a fentiekben ontológiai / szemantikai szelekciós információk nélkül nem dönthető el, hogy a *szög* a *szűrés* eszköze vagy ágense-e: a teszteredményekben – nagyon helyesen – mindkét értelmezést viszontláthatjuk. A két értelmezés között csak egy szemantikus információt is tartalmazó lexikonszerkezet tud egyértelműen választani: az ágensnek önálló cselekedetre képes (pl. az *Agent* fogalomból leszármaztatható) objektumnak kell lennie.

5 Összegzés, értékelés

Miután az 1. szakaszban bemutattuk a *ReALIS* dinamikus diskurzussemantikai interpretációs modellt [4], majd a 2.-ban a polaritásos hatásláncsaládok lexikai-szemantikai elméletét [2, 3], a 3.-ban pedig az utóbbi beépítését az előbbibe a σ függvény paramétervektora révén, a 4. szakaszban egy implementáció sarokköveit vázoltuk fel, részeredményekről számolva be. Eddig alapelvek és adatformátumok rögzítése és tanulmányprogramok írása történt meg. A helyzetet egyrészt bonyolítja, másrészt megkönnyíti a kutatás korábbi szakaszában létrejött szóalak-elemzési program [5, 6] integrálásának igénye. Egy teljes *ReALIS*-megvalósításról egyelőre korai még a beszámoló, de ilyen irányba haladunk, és ez további közlemények tárgya lesz.

Bibliográfia

1. Alberti, G.: Restrictions on the Degree of Referentiality of Arguments in Hungarian Sentences. In: *Acta Linguistica Hungarica* Vol. 44 No. 3–4 (1997) 341–362
2. Alberti, G.: A szóképzéssel együttjáró vonzatszerkezet-változások rendszere. In: *Nyelvtudományi Közlemények* No. 103 (2006) 75–105
3. Alberti, G.: A szóképzéssel együttjáró vonzatszerkezet-változások egy polaritásérzékeny rendszere. In: Fancsaly É. (szerk.): *Tanár és tanítvány. Írások Györke József és Hajdú Péter tiszteletére* (2002-2007). *Studia Linguistica, Dialóg Campus*, Bp. – Pécs (2009) 122–145
4. Alberti, G.: *ReALIS*, avagy a szintaxis dekompozíciója. Általános Nyelvészeti Tanulmányok. Megjelenés előtt.

5. Alberti, G., Kleiber J.: The *GeLexi* MT Project. In: Hutchins, J. (szerk.): Proceedings of EAMT 2004 Workshop (Malta). Univ. of Malta, Valletta (2004) 1–10
6. Alberti, G., Kleiber J.: The Grammar of ReALIS and the Implementation of its Dynamic Interpretation. *Informatica* Vol. 34 No. 2 (2010) 103–110
7. Alberti, G., Károly, M., Kleiber, J.: The ReALIS Model of Human Interpreters and Its Application in Comp. Ling. In: Proc. ICSOFT 2010/2. SciTePress Portugal (2010) 468–474
8. Asher, N., Lascarides, A.: Logics of Conversation. Cambridge Univ. Press (2003)
9. de Boer, T. W.: A Beginners' Guide to Visual Prolog. Prolog Development Center A/S, Copenhagen, Denmark (2009)
<http://download.pdc.dk/vip/72/books/deBoer/VisualPrologBeginners.pdf>
10. Croft, W.: Radical Construction Grammar. Syntactic Theory in Typological Perspective. Oxford University Press (2001)
11. Dowty, D. R., Wall, R. E., Peters, S.: Introduction to Montague Semantics. D. Reidel Publishing Company, Dordrecht (1981)
12. van Eijck, J., Kamp, H.: Representing discourse in context. In: vBenthem, J., ter Meulen, A. (szerk.): Handbook of Logic and Language. Elsevier, Amsterdam (1997) 179–237
13. Kilián, I.: Horn clauses. A Two Way Street. Kézirat: Gyűrűfü Műhely Kft.. (korábban publikálva a www.sics.se honlapon) (1996)
14. Kiss, É. K.: Többszörös fókusz a magyar mondat szerkezetben. In: Büky, L., Maleczki, M. (szerk.): A mai magyar nyelv leírásának újabb módszerei II. (1995) 47–66
15. Komlósy, A.: Régensek és vonzatok. Kiefer, F. (szerk.): Strukturális magyar nyelvtan. I. Mondattan. Akadémiai Kiadó, Budapest (1992) 299–527
16. Nakashima, H.: Term Description: A Simple Powerful Extension to Prolog Data Structures. Electrotechnical Laboratory, Umezono, 1-1-4, Ibaraki, Japan (1985)
17. Pléh, Cs.: Topic and subject prominence in Hungarian. In Kiefer, F. (szerk.): Hungarian Linguistics (Linguistic and Literary Studies in Eastern Europe Vol. 4). John Benjamins, Amsterdam (1982)